Skew Angle Detection of Bangla Script using Radon Transform

S. M. Murtoza Habib, Nawsher Ahamed Noor[†] and Mumit Khan
Center for Research on Bangla Language Processing, BRAC University, Dhaka, Bangladesh.
[†]Department of Computer Science and Engineering, BRAC University, Dhaka, Bangladesh.
murtoza@gmail.com, nawsher@gmail.com, mumit@bracu.ac.bd

Abstract

Skew angle detection and correction an integral part of any OCR system. Without proper skew correction, the performance of an OCR will simply not be acceptable for most scanned images. We propose an innovative method for skew angle detection and correction for Bangla scripts using the Radon Transform. The basic idea is to identify the upper envelope by detecting the headline that accompanies most of the letters in the Bangla script, and then apply the Radon Transform to this upper envelope to get the skew angle. Once the angle is known, the correction is quite trivial to perform. While the current implementation handles only a single skew angle per text image, it can be extended to handle multiple skew angles by partitioning the document image.

Keywords: upper envelope, skew angle, radon transform.

I. INTRODUCTION

Optical Character Recognition (OCR) is a technology to convert digital image of text to editable text. A typical usage of an OCR is to convert a scanned image of text on paper to a text document that can be further processed on a computer. Before the actual character recognition step, the input image must be pre-processed; the first step is to recognize the regions of the image as text, and only apply recognition on those regions; the image is then corrected for skewing, with each region of text re-aligned before going on to the next stage; then the lines, words and perhaps even characters are segmented to form the minimal constituents that are then recognized by the OCR. When a document is fed to the optical sensor (scanner) to get the digital image, there may be a few degrees of skew. Skew angle is the angle that the text lines in the digital image make with the horizontal direction. Therefore, skew estimation and correction are important steps before line separation. This paper deals with skew angle detection and correction.

A recent work in this area uses the Digital Straight Line (DSL) to find the skew angle in the scanned images of text [1]. This DSL-based method provides a good compromise between performance and accuracy; in comparison, our Radon Transform-based method is more accurate, but the runtime performance is slightly slower than [1].

We review the properties of the properties of the Bangla script and the theory of Radon Transform in sections II and III; we then describe our methodology and evaluate our implementation in section IV; finally, we conclude with a look at future work in section V.

II. SOME PROPERTIES BANGLA SCRIPT

The writing style of Bangla is from left to right with 11 independent vowel and 39 consonant characters. We may be called these characters as basic characters (Fig. 1). The concept of upper/lower case is absent in Bangla. From Fig. 1 it is noted that most of the characters have a horizontal line at the upper part. This horizontal line is called headline. In Bangla language, we call this line 'matra'.

Out of these 11 independent vowels 10 vowels has dependent form. The dependent vowels with examples are shown in Table 1. If the first character of the word is a vowel then it remains its independent form. Generally a vowel following by a consonant takes a dependent form and placed at the left or right or both or bottom of the consonant.

অ আ ই ঈ উ ঊ
ঋ এ ঐ ও ঔ

(a) Vowels

ক খ গ ঘ ঙ চ ছ জ ঝ ঞ ট ঠ ড
ঢ ণ ত থ দ ধ ন প ফ ব ভ ম য র
ল শ ষ স হ ড় ঢ় য় র ৎ ◌ং ◌ঃ ◌ঁ

(b) Consonants

Figure 1. (a) Vowels; (b) Consonants

For two consecutive vowels following by a consonant in a word, the second one remains in its independent form as shown Fig. 2.

খ + আ + ও -> খাও

Figure 2. A word with dependent and independent vowels

A consonant followed by a consonant sometimes takes a compound shape which we call as compound character as shown in Table 2. There are about 250 compound characters, most of which are formed by consonant-consonant combination. Compounding of three consonants is also possible. Most interesting thing is if we change the order of two consonant the compound character is changed.

Table 1. Example of Independent and dependant vowels

	Independent Shape	Example	Dependent shape	Example
1	আ	আমরা	◌া	কাল
2	ই	ইদুর	◌ি	তিন
3	ঈ	ঈদ	◌ী	নীট
4	উ	উট	◌ু	তুমি
5	ঊ	ঊষা	◌ূ	পূজা
6	ঋ	ঋণ	◌ৃ	কৃষি
7	এ	এক	◌ে	কেমন
8	ঐ	ঐরাবত	◌ৈ	তৈল
9	ও	ওল	◌ো	তোমা
10	ঔ	ঔষধ	◌ৌ	নৌকা

Table 2. Some compound characters

Consonant combination	Compound character
ক ◌্ ক	ক্ক
ক ◌্ ত	ক্ত
ক ◌্ ম	ক্ম
ঙ ◌্ ক	ঙ্ক
জ ◌্ ঞ	জ্ঞ
ঞ ◌্ জ	ঞ্জ
ক ◌্ ষ ◌্ ম	ক্ষ্ম

As we mentioned earlier, most of the basic characters have a head line (matra). Out of 50 basic characters 32 characters have this matra (head line). Here we can detect the skew angle using this matra.

III. THE RADON TRANSFORM

In recent years the Hough transform and the related Radon transform have received much attention. These two transforms are able to transform two dimensional images with lines into a domain of possible line parameters, where each line in the image will give a peak positioned at the corresponding line parameters. This has lead to many line detection applications within image processing, computer vision, and seismic.

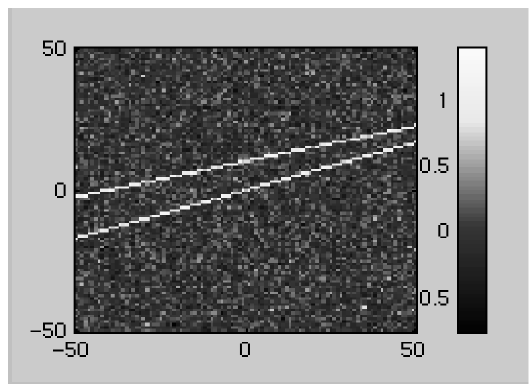
Several definitions of the Radon transform exists, but they are related, and a very popular form expresses lines in the form $\rho = x\cos\theta + y\sin\theta$, where θ is the angle and ρ the smallest distance to the origin of the coordinate system. As shown in the two following definitions (which are identical), the Radon transform for a set of parameters (ρ,θ) is the line integral through the image g(x, y), where the line is positioned corresponding to the value of (ρ,θ). The $\delta(\ )$ is the delta function which is infinite for argument 0 and zero for all other arguments (it integrates to one).

$$\breve{g}(\rho,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\delta(\rho - x\cos\theta - y\sin\theta)dx\ dy$$

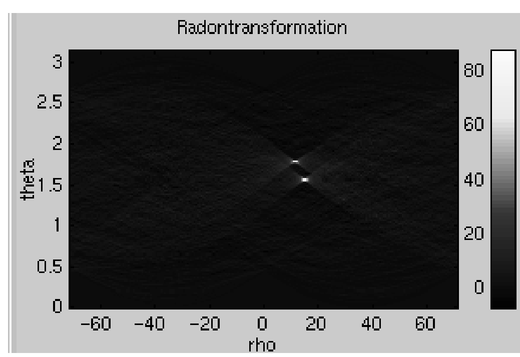
or the identical expression

$$\breve{g}(\rho,\theta) = \int_{-\infty}^{\infty} g(\rho\ \ \cos\theta - s\ \ \sin\theta,\ \ \rho\ \ \sin\theta + s\ \ \cos\theta)ds$$

Using this definition an image containing two lines are transformed into the Radon transform shown to the right



a) Input image



a) Transform result

Figure 3. Example of Radon Transform

It can be seen that two very bright spots are found in the Radon transform, and the positions shown the parameters of the lines in the original image. A simple threshold algorithm could then be used to pick out the line parameters, and given that the transform is linear many lines will just give rise to a set of distinct point in the Radon domain. In his Ph.D. thesis [3], Peter Toft investigated the relationship of Radon transform with the Hough transform, and it is shown that the Radon transform and the Hough transform are related but not the same.

IV. METHODOLOGY AND RESULT

From [1] and [2] we found some important statistics for Bangla language. Some important statistics are:

1. The average length of Bangla words is about six characters.
2. About 30%-35% of characters are vowel modifiers which, being small in size, contribute very little to the head line of the word.
3. Most basic characters are consonants, as vowels in basic form can appear at the beginning of the word or when two vowels appear side by side.
4. Compound characters are very infrequent, occurring in about 5% of the cases only.
5. In Bangla 41 characters can appear in the first position of a word. Out of these 41 characters 30 characters have head lines.
6. Probability of getting a character with head line in the first position of a word is $p_1 = \frac{30}{41}$ and getting a character without head line in the first position is $1 - p_1 = \frac{11}{41}$.
7. In other positions of a word are mostly consonants and 28 out of 39 Bangla consonants have head lines.
8. Probability of getting a consonant with head line for other positions in a word is $p_2 = \frac{28}{39}$ and probability of getting a character without head line in other positions is $1 - p_2 = \frac{11}{39}$.
9. Thus, probability of all four characters without head line in a word is $(1 - P_1)(1 - P_2)^3 = 0.00601$ (assuming that all characters are equally likely and independently occurring in a word). Hence, probability that a word will have at least one character with head line is 1 - 0.00601 = 0.99399. The practical situation is better than these estimates since characters are not equally likely in a word and most frequently used characters have head lines.

In Bangla, a head line connects all characters in a word; therefore we can detect a word by the method of connected component labeling. As mentioned in [1], for skew angle detection, at first the connected component labeling is done. At the time of component labeling, for each labeled component its bounding box (minimum upright rectangle containing the component) is defined. The mean b_m and standard deviation b_s of the boundary box width are also computed. Next, components having boundary box width greater than or equal to b_m and less than $b_m + 3b_s$ are preserved. By threshold at b_m the small components like dots, punctuation marks, isolated characters and characters without head line are mostly filtered out while by threshold at $b_m + 3b_s$ big components that may represent graphs and tables are also filtered out. Because of these filtering processes the irrelevant components cannot create error in skew estimation.

Then we find *upper envelope* of the selected components as mentioned [1]. From each pixel of the uppermost row of a bounding box, we perform a vertical scan and as soon as a pixel is found. The set of pixels obtained in this way denotes the upper envelope of the component. In Fig. 4 we take digital image of text having 30 degree of skewed angle. In Fig. 5 we show the Upper Envelope of Fig. 4 after threshold as previous. Note that in most of the cases the upper envelope contains the headline. In this way we could filter out the irrelevant data for further processing. Here we will apply Radon Transform technique on the upper envelopes for skew estimation.

Figure 4. Digital image of text having 30 degree of skewed angle.

The Radon transform of a function f(x,y) is defined as the integral along a straight line defined by its distance P from the origin and its angle of inclination θ, a definition very close to that of the Hough transform and requires a lot of processing power in order to be able to do its work in a reasonably finite time. Now a day high processing power is not a problem. All processor speed in the market is now more than one GHz and main memory also very cheap. So, to use Radon Transform will not be a problem.

The Radon transform can detect a line in any angle.

Here we are considering all the line has same skew angle and the range of angle is -45^{o} to 45^{o}. Here Radon transform will detect the angle from the upper envelope. If the skewed angle is more than 45^{o} or less then -45^{o} the upper envelope may contain 2 lines in different direction. An example is shown in Fig. 6 having 50 degree of skewed angle which may create a problem.

After detect a lines Radon transform give the angles and distance of lines from the origin. The Radon transform of Upper Envelope (Fig. 5) is shown in Fig. 7.

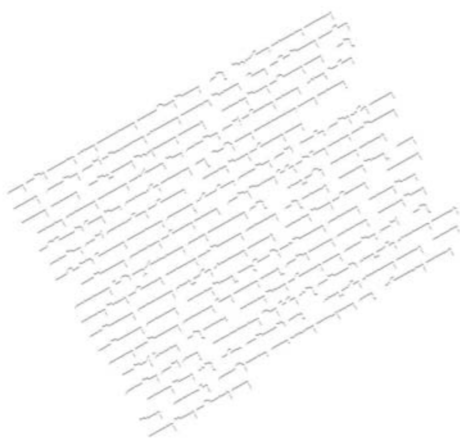


Fig. 5 Upper Envelope of Fig. 4 after threshold

Figure 6. Two straight lines in a single word

In Fig. 7, the point which has intensity high is repeating a straight line. Here we can see that all the straight line has the same degree which is 120. To find the desired angle we have to subtract this angle from 90 degree and the desired skewed angle is 90 – 120 = –30. So, if we rotate the digital image (Fig. 4) in degree –30 then we will find our desired image (Fig. 8).

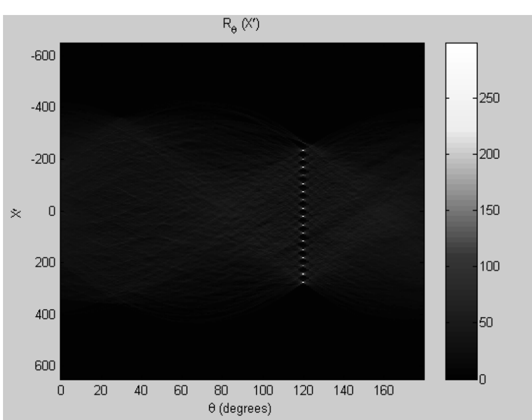


Figure 7. Radon transform of Upper Envelope (Fig. 5)

In this case, the accuracy of Radon Transform is 100%. Table 3 shows that, for a wide range of skew angles, the algorithm detects the correct angle in each case.

Figure 8. Image after rotate

Table 3 Performance of Radon Transform based skew angle detector

Original skew angle in degree(manual)	Skew angle using propose method
35	35
30	30
20	20
10	10
5	5
2	2

V. CONCLUSION

Skew detection is a fundamental step in the preprocessing stage of any OCR system. The input must be properly aligned with respect to some axis before the recognition engine can achieve the high level of accuracy, often approaching 100%, offered by today's commercial OCR systems. The complexity of the Brahmi family of scripts poses a challenge when computing the skew angle. We propose an innovative method of computing the skew angle using the Radon Transform, often achieving 100% accuracy over a wide range of skew angle (±45 degrees). Future extensions to the current implementation will handle multi-skew document images by partitioning the image and then performing this algorithm on each of the partitions.

VI. ACKNOWLEDGEMENT

This work has been supported in part by the PAN Localization Project (www.panl10n.net) grant from the International Development Research Center, Ottawa, Canada, administrated through Center for Research in Urdu Language Processing, National University of Computer and Emerging Sciences, Pakistan.

REFERENCES

[1] B.B. Chaudhuri and U. Pal, “*Skew Angle Detection of Digitized Indian Script Documents*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 2, February 1997.

[2] B.B. Chaudhuri and U. Pal, “*Relational Studies Between Phoneme and Grapheme Statistics in Modern Bangla Language*”, J. Acoustical Society of India, vol. 23, pp. 67-77, 1995.

[3] Peter Toft: "*The Radon Transform - Theory and Implementation*", Ph.D. thesis. Department of Mathematical Modelling, Technical University of Denmark, June 1996. 326 pages.